Assessing skin lesion evolution from multispectral image sequences

Sylvain Prigent, Xavier Descombes, Didier Zugaj, Laurent Petit, Anne-Sophie Dugaret, Philippe Martel, Josiane Zerubia

Assessing skin lesion evolution from multispectral image sequences

Sylvain Prigent, Xavier Descombes, Didier Zugaj, Laurent Petit , Anne-Sophie Dugaret , Philippe Martel , Josiane Zerubia

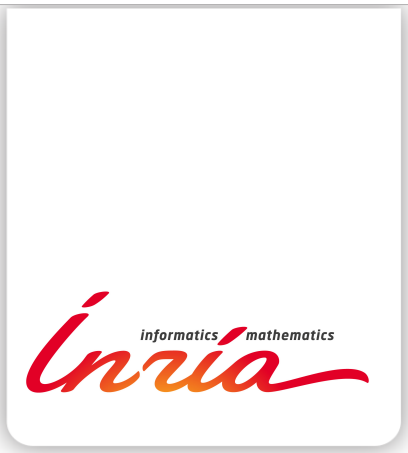
RESEARCH REPORT

N° 8745

June 2015

Project-Teams Morphème and Ayin

ISSN 0249-6399 ISRN INRIA/RR--8745--FR+ENG



Assessing skin lesion evolution from multispectral image sequences

Sylvain Prigent *, Xavier Descombes†, Didier Zugaj‡,
Laurent Petit ‡, Anne-Sophie Dugaret ‡, Philippe Martel ‡,
Josiane Zerubia §

Project-Teams Morphème and Ayin

Research Report n° 8745 — June 2015 — 26 pages

Abstract: During the evaluation of skin disease treatments, dermatologists have to clinically measure the evolution of the pathology severity of each patient during treatment periods. Such a process is sensitive to intra- and inter- dermatologist diagnosis. To make this severity measurement more objective we quantify the pathology severity using a new image processing based method. We focus on a hyperpigmentation disorder called melasma. During a treatment period, multispectral images are taken on patients receiving the same treatment. After co-registration and segmentation steps, we propose an algorithm to measure the intensity, the size and the homogeneity evolution of the pathological areas. Obtained results are compared with a dermatologist diagnosis using statistical tests on two clinical studies containing respectively 384 images from 16 patients and 352 images from 22 patients.
This research report is an update of the report 8136 [49]. It describes methods and experiments in more details and provides more references.

Key-words: skin, hyperpigmentation, change detection, statistical inference, multispectral

* Morphème/Ayin
† Morphème
‡ Galderma R&D
§ Ayin

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Quantification de l'évolution de lésions cutanées par imagerie multispectrale

Résumé : Lors de l'évaluation des traitements des maladies de peau, les dermatologues doivent mesurer la sévérité de la pathologie de chaque patient tout au long d'une période de traitement. Un tel procédé est sensible aux variations intra- et inter- dermatologues. Pour rendre cette mesure de sévérité plus robuste, nous proposons d'utiliser l'imagerie spectrale. Nous nous concentrons sur une pathologie d'hyperpigmentation cutanée appelée mélasma. Au cours d'une période de traitement, des images multispectrales sont acquises sur une population de patients sous traitement. Après des étapes de recalage des séries temporelles d'images et de classication des régions d'intérêt, nous proposons une méthodologie permettant de mesurer, dans le temps, la variation de contraste, de surface et d'homogénéité de la zone pathologique pour chaque patient. Les résultats obtenus sont comparés à un diagnostique clinique à l'aide de tests statistiques réalisés sur une étude clinique complète.

Ce rapport de recherche est un complément du rapport de recherche 8136 [49], afin de compléter la bibliographie, et de décrire plus en détail les méthodes et résultats.

Mots-clés : peau, hyperpigmentation, d'étection de changements, inférence statistique, multispectral

Contents

1 Introduction

In dermatology, specific protocols are established to quantify the evolution of lesion severity along a treatment period. In this report, we focus on a hyperpigmentation disorder called melasma. This disease produces darker and irregular spots on the face. Melasma has the specificity to be symmetric w.r.t the two cheeks. Thereby, the protocol consists in applying a studied treatment on one cheek, and a comparator treatment on the second cheek ([40]). The comparator is often chosen as the medium used for the studied treatment without active product. To render a clinical study reliable, patients are randomized ([40]). The efficacy of a treatment is then clinically measured by dermatologists. To this end, a clinical score is evaluated by one or several dermatologists on each cheek of each patient during several visits along the treatment. Finally, a statistical test such as the Student test or the Wilcoxon test ([36]) is performed on these measurements to obtain an evaluation of the studied treatment efficacy relatively to the comparator treatment ([41]). The most widely used score for melasma severity evaluation is the "Melasma Area and Severity Index" (MASI) ([34, 21, 3]). It has been introduced in ([26]). The MASI splits the face into 4 regions: forehead, left cheek, right cheek, and chin. For each region, three scores are evaluated:

- *Area (A):* Measure of the area affected by the disease. It is estimated on a scale ranging from 0 to 6 as: $0 = 0\%$; $1 \leq 10\%$; $2 = 10-29\%$; $3 = 30-49\%$; $4 = 50-69\%$; $5 = 70-89\%$ and $6 = 90-100\%$.
- *Darkness (D):* Color of the pathological area compared with normal skin. It is graded on a scale from 0 to 4 as follows: 0=normal skin color without evidence of hyperpigmentation; 1=barely visible hyperpigmentation; 2=mild hyperpigmentation; 3=moderate hyperpigmentation; 4=severe hyperpigmentation.
- *Homogeneity (H):* Homogeneity of the hyperpigmentation, graded on a scale from 0 to 4 as follows: 0=normal skin color without evidence of hyperpigmentation; 1=specks of involvement; 2=small patchy areas of involvement < 1.5 cm diameter; 3=patches of involvement > 2 cm diameter; 4=uniform skin involvement without any clear area).

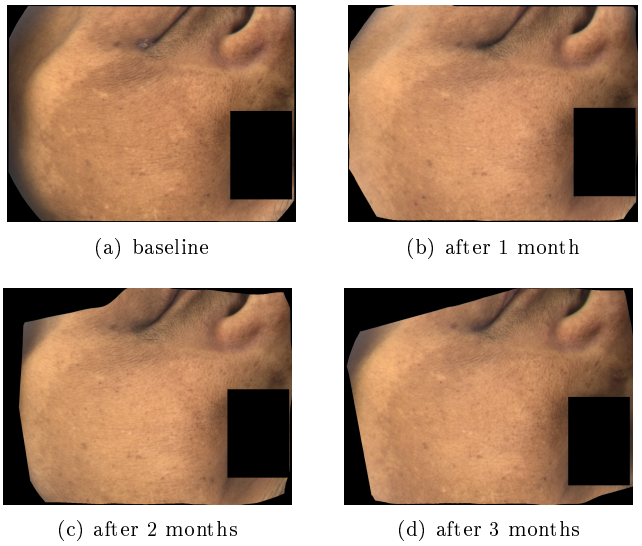
To integrate these three criteria measured on the four regions: forehead (F), left cheek (LC), right cheek (RC) and chin (C), the following equation is proposed in [26]:

$$MASI = 0.3(D_F + H_F)A_F + 0.3(D_{LC} + H_{LC})A_{LC} + 0.3(D_{RC} + H_{RC})A_{RC} + 0.1(D_C + H_C)A_C, \quad (1)$$

where D_F, D_{LC}, D_{RC}, D_C are darknesses measured respectively on the forehead, left cheek, right cheek and chin, H_F, H_{LC}, H_{RC}, H_C homogeneities measured respectively on the forehead, left cheek, right cheek and chin and A_F, A_{LC}, A_{RC}, A_C areas measured respectively on the forehead, left cheek, right cheek and chin. This MASI definition has been discussed in [41]. The authors claim that the H criterion has significant variations from one dermatologist to another. They then propose to remove the H criterion from Equation (1):

$$MASI = 0.3D_FA_F + 0.3D_{LC}A_{LC} + 0.3D_{RC}A_{RC} + 0.1D_CA_C. \quad (2)$$

In this report, we propose a new computer based score called "differential score" that measures the variation of the pathology between two visits. The idea is to consider multispectral images from which we automatically compute a severity score containing area (A), darkness (D) criteria as defined by the MASI. As the homogeneity (H) criterion is expert dependent, we introduce a new



(a) baseline (b) after 1 month

(c) after 2 months (d) after 3 months

Figure 1: Time series of color images for a patient whose pathology was diagnosed as decreasing. The diagnosis of the dermatologist using a scale from 0 to 3 was 2 (moderate) at baseline and 0 (cleared) after 3 months.

criterion H which measures the local variations of the pathology. Such a computer based score avoids the variability on the measures obtained by human inspection. The choice of multispectral imaging is motivated by the fact that we need more information than color imaging to evaluate the Melasma evolution. As an example, Figure 1 shows a time series of co-registered color images. The diagnosis of the dermatologist is a decreasing of the Melasma over time. Looking at color images of Figure 1 doesn't allow to easily visualize such a decreasing.

The report is organized as follows: section 2 introduces the proposed methodology and the chosen algorithms to extract the disease information from multispectral images and define the "differential score". In section 3, we present the experimental settings: we describe the two clinical studies that we work with and expose the protocol used to evaluate the proposed "differential score". In section 4 we expose the obtained results using the two clinical studies. Finally, section 5 discusses the proposed score and concludes this report.

2 Method

In this section, we introduce the methodology to score the disease evolution with a time series of multispectral images. After exposing the global scheme and the state of the art, we describe the selected algorithms for extracting the information needed to compute the "differential score" from images.

2.1 Proposed approach and state of the art

The quantification of a treatment efficacy along a clinical study using imaging is the comparison of the disease severity along a sequence of images for each patient. To quantify this severity, we first need to extract the pathological areas in the images, and then extract features describing the disease. Furthermore, we aim at measuring changes of the features from time t_0 to time t_k.

Figure 2 summarizes the proposed approach. In this scheme, a multispectral (MS) image has been taken for each patient at different periods ("MS images t_0", "MS images t_k", "MS images t_n"). We first segment each image taken at baseline (t_0) in order to extract the pathological area that is the region of interest (ROI) where the disease severity evolution along the clinical study is measured. To be able to get the same ROI for each patient at different times, we co-register the sequence of images of each patient individually. This co-registration is done using the algorithm described in [37] initialized by an affine registration.

To evaluate the severity, we first extract a spectral feature that provides a local severity indicator on each pixel of the image. we then calculate the severity evolution using some change detection approaches. A binary change detection is used to measure a global evolution of the disease, and a multi-level change detection is used to evaluate the local temporal evolution.

Figure 2: Proposed scheme to evaluate a treatment efficacy on a clinical study

Several methods have been described to perform the 3 steps (segmentation, feature extraction and change detection) above mentioned:

Segmentation Numerous methods have been proposed to classify or segment multispectral and hyperspectral images. Since the spectral space is large, a dimension reduction is usually performed to avoid the Hughes phenomenon ([20]). Classification or segmentation methods can then be partitioned into three groups.

The first group is only based on spectral features (for instance kmeans [[24]), fuzzy-C-means ([6]), mixture analysis ([10]), mean shift ([12]), k-nearest neighbors ([13]), maximum likelihood ([35]) or SVM ([62])]. The second group consists of the methods that extend the gray scales image algorithms to spectral images [Markov random fields ([19, 51]), extension of split and merge algorithms ([25, 65, 59, 60]), and extension of morphological mathematics ([43, 5])]. Finally we can find some combination of the two previous groups [([5, 61, 58])].

In this report, we propose to extract the relevant information using a projection pursuit algorithm, to remove the shading artifacts using a compensation and then to perform the segmentation using a SVM that embeds non linear separable classes ([48]).

Feature extraction Several approaches have been proposed to analyse the spectrum of the skin from multipsectral images. The most popular spectral analysis algorithm for the skin is probably the one proposed in [56, 57]. The authors model the skin absorbency as a mixture of melanin, deoxy-haemoglobin and oxy-haemoglobin components. A linear regression associated with the Beer-Lamber law provides the proportions of these three components at each pixel. More generally, feature extraction for spectral images has been widely studied. Methods of features extraction considering a linear combination of spectral bands are PCA ([16]), NAPCA ([18]), SVD ([4]), or ICA ([9, 22]) that tend to extract features by minimizing the correlation or higher order statistics. The NMF approach ([39]) searches for features by factorizing the signal matrix into two non negative matrices. This factorization can be done by minimizing the mean square error or the Kullback-Leibler divergence for example ([14, 30]). Algorithms such as N-FINDR ([63, 15, 11]), PPI ([7, 42]) or VCA ([38]) analyse the geometry of the data set in the spectral N_b dimensional space to extract features defined as extreme points in this set. We can also cite supervised methods like DAFE ([27]), NWFE ([28, 29]) or a maximum likelihood based approach ([52]).

In this report, to reduce the computational time induced by the huge volume of multispectral images, we propose to select the "best" band with respect to the contrast criterion. We have shown in [47] that although simple, this approach provides results comparable to more sophisticated techniques such as ICA.

Change detection In the literature, numerous methods have been proposed to quantify the changes between two images. They consider the image either globally using pixelwise operations such as subtraction or division ([33, 50, 54, 32, 31]) or locally with a window that browses the image ([45, 1, 53, 8, 23, 64]). In each cases, a change map is thresholded to obtain a binary map of changes.

To take into account spatial information we adapt in this report the SPM approach developed to analyze fMRI data ([17]). This last step represents our main contribution that leads to a differential score quantifying the efficiency of a treatment.

2.2 Segmentation

The goal of the segmentation step is to select a region of interest in each image at baseline (t_0) to be followed during the time sequence. In this paragraph, we summarize a work detailed in [46] and [48]. The segmentation of skin hyperpigmented areas is performed in three steps. First, the image spectrum dimensionality is reduced in order to remove redundancy and avoid the Hughes phenomenon using a projection pursuit algorithm ([46]). Second, the shading effect due to the volumes of the face (nose, eyes) (see Figures 3(a), 3(d) and 3(g)) is compensated to ensure an accurate segmentation. We first remove the background inhomogeneity using a Fourier high pass filter and subtract the near infrared spectral band from each of the other bands. The near infrared has been chosen to remove shading since this band contains few skin pigmentation information. On this transformed dataset, the segmentation is performed using support vector machine (SVM). The SVM is trained on a few representative normalized images of the clinical study. The SVM can then generalize to segment all the images. Examples of training data are shown in Figures 3(b), 3(e) and 3(h). The variability of the disease between patients makes this segmentation task difficult. To be sure the pathological area is detected on each image, we created an interactive software that allows a human operator to manually correct the segmentation if necessary. We generate two other segmentations by Gaussian mixture analysis associated with the connected component analysis using two sets of parameters ([48]). We also partition the SVM segmentation using the connected component analysis. We then get three segmentation. These maps allow an operator to select the area of interest by clicking the components to be kept. Figures 3(c), 3(f) and 3(i) show the obtained segmentation on three different patients. Figure 4 shows an example of correction of the segmentation using the two additional segmentation.

2.3 Spectral criterion: contrast

To extract a spectral feature of the studied disease, we assume that we have extracted a region of healthy skin and a region of pathological skin for each series of co-registered multispectral images. For a given patient, at a given time t_k, we thus have four regions: $I_{t_k}^{pS}$ and $I_{t_k}^{hS}$ the pathological and healthy regions on the cheek treated with the studied treatment, and $I_{t_k}^{pC}$ and $I_{t_k}^{hC}$ the pathological and healthy regions on the cheek treated with the comparator treatment. We consider these four regions to compute the weights of a linear combination M of the spectral bands:

$$M(I) = \sum_{b=1}^{N_b} m_b I(b), \tag{3}$$

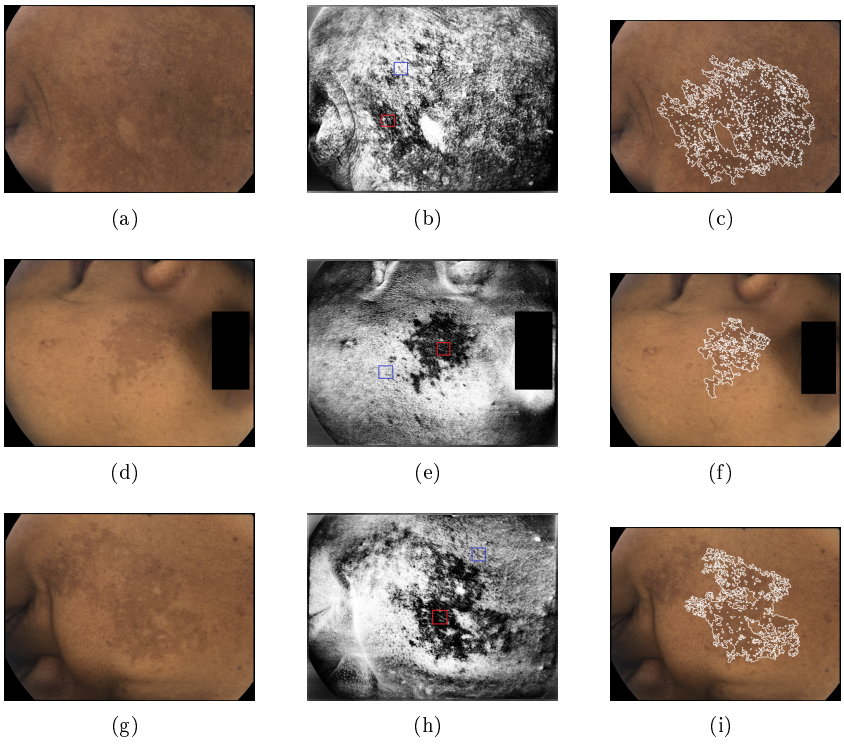


Figure 3: Images used for the global training. (a,d,g) are color reconstructed images by spectral integration. The color reconstruction is used only for visualization. (b,e,h) are filtered images by volume compensation and Fourier high pass filter at 40%, red rectangles show training pathological areas, and blue rectangles show training healthy areas. (c,f,i) are the surrounding of the pathological area detected by the proposed method.

where m_b is the weight of the b^{st} spectral band $I(b)$ of the image I and N_b the number of spectral bands in the initial multispectral image. We look for the combination that gives the highest contrast between the healthy areas and the pathological areas.

We first introduce a normalization step that combines the four regions $I_{t_k}^{pS}$, $I_{t_k}^{hS}$, $I_{t_k}^{pC}$, $I_{t_k}^{hC}$. The first normalization combines the healthy and the pathological areas. We can either consider only the pathological area or compute the difference between healthy and pathological areas. The second normalization combines the studied treatment S and the comparator treatment C. We use the method proposed in clinical protocols that calculates the difference $S-C$. Depending on the use or not of the healthy area, $S = \mu_M(I_{t_k}^{e,pS})$ or $S = \mu_M(I_{t_k}^{e,hS}) - \mu_M(I_{t_k}^{e,pS})$ and $C = \mu_M(I_{t_k}^{e,pC})$ or $C = \mu_M(I_{t_k}^{e,hC}) - \mu_M(I_{t_k}^{e,pC})$, we have two possible normalizations:

$$D_{t_k}^e = \mu_M(I_{t_k}^{e,pS}) - \mu_M(I_{t_k}^{e,pC}) \tag{4}$$

and

$$D_{t_k}^e = (\mu_M(I_{t_k}^{e,hS}) - \mu_M(I_{t_k}^{e,pS})) - (\mu_M(I_{t_k}^{e,hC}) - \mu_M(I_{t_k}^{e,pC})), \tag{5}$$

where $D_{t_k}^e$ denotes the darkness for the patient e at time t_k and $\mu_M(I_{t_k}^{hS})$ (respectively $\mu_M(I_{t_k}^{e,pS})$, $\mu_M(I_{t_k}^{e,hC})$, $\mu_M(I_{t_k}^{e,pC})$) is the spatial average of the spectral feature M over the area $I_{t_k}^{e,hS}$ (respectively $I_{t_k}^{e,pS}$, $I_{t_k}^{e,hC}$, $I_{t_k}^{e,pC}$).

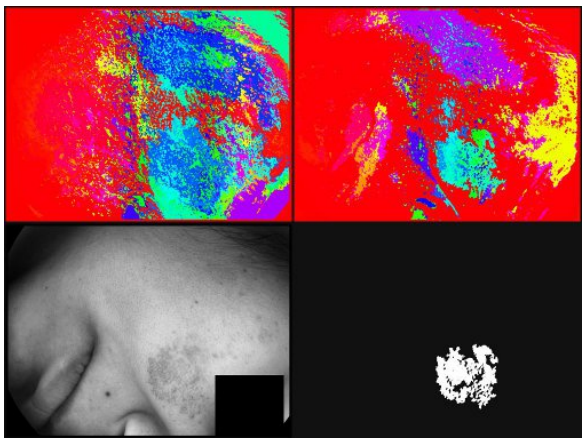


Figure 4: Screenshot of the software that allows to do the interactive segmentation, first example. Top left: segmentation obtained with a first set of parameters, top right: segmentation obtained with a second set of parameters, bottom left one band of the original image, bottom right SVM chosen component. The user can add a region in the top row to the segmentation obtained from SVM.

We propose an objective function that maximizes both the distance between pathological and healthy areas, and the deviation between measurement times. This objective function can be written as $f = \sum_{t=t_1}^{t_n} \sum_{e=1}^{N_e} D^e_{t_k} - D^e_{t_0}$. Using $D^e_{t_k}$ of equation (5) we get:

$$f = \sum_{b=1}^{N_b} m_b \left\{ \sum_{t=t_1}^{t_n} \sum_{e=1}^{N_e} \Big[\mu_{I^{e,hS}_{t_0}}(b) - \mu_{I^{e,pS}_{t_0}}(b) \; - \mu_{I^{e,hC}_{t_0}}(b) + \mu_{I^{e,pC}_{t_0}}(b) \right.$$
$$\left. - \mu_{I^{e,hS}_{t}}(b) + \mu_{I^{e,pS}_{t}}(b) \; + \mu_{I^{e,hC}_{t}}(b) - \mu_{I^{epC}_{t}}(b) \Big] \right\}. \quad (6)$$

The function f is linear with respect to the parameters m_b of M. The optimization is then trivial. The solution is the spectral band with the highest contrast if we consider only positive m_b. If we consider also negative m_b, the solution is given by two bands.

The darkness criterion is then defined by $D^e_{t_k}$ from equation (4) or (5) depending on the chosen normalization, and using the equation (6) to define the linear combination M.

2.4 Change detection

In this section, we describe the two selected methods that we use to extract a binary change map and a multi-level change map. The binary change map aim at measuring the global area of changes whereas the multi-level change map intends to evaluate the local changes.

2.4.1 Binary change detection: area

Averaging on a local window to quantify the changes according to each pixel neighborhood tends to a loss in spatial accuracy that is critical. We perform change detection by subtraction to avoid the noise induced by division:

$$F^e_{t_{k/0}} = M(I^e_{t_k}) - M(I^e_{t_0}), \quad (7)$$

where $M(I^e_{t_k})$ (respectively $M(I^e_{t_0})$) is the spectral feature M of $I^e_{t_k}$ (respectively $I^e_{t_0}$) defined in equation (3). A threshold τ on $F^e_{t_{k/0}}$ allows to get the binary change map:

$$B^e_{t_{k/0}}(x) = \begin{cases} 1 & \text{if } F^e_{t_{k/0}}(x) > \tau \text{ and } C^e_{t_0}(x) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where x denotes the position of any pixel in the image. The choice of τ is crucial. We propose a method that computes one threshold τ per image series estimated on the histogram of $F^e_{t_{1/0}}$. First, we assume that the histogram of $F^e_{t_{k/0}}$ is monomodal, symmetric and centred in zero in the case of no changes. This assumption is justified since $F^e_{t_{k/0}}$ is a centred noise in the case of no changes. Then, we assume that there are no significant changes in the negative part (i.e. the disease does not become worst). Thus, in the case of no disease decreasing, the number of positive and negative pixels in $F^e_{t_{k/0}}$ should statistically be the same.

The threshold τ is calculated such that: $\text{card}\{x : 0 < F^e_{t_{k/0}}(x) < \tau\} = \text{card}\{x : F^e_{t_{k/0}}(x) \leq 0\}$.

A drawback of the thresholding methods for change detection is their sensitivity to noise. To decrease the noise, we filter the change map $B^e_{t_{k/0}}$ with opening and closing from mathematical morphology operators ([55]) using a disk of two pixels as a structural element.

To calculate an area criterion, we propose to compute the spatial changes between t_k and t_0. Let us denote $C^e_{t_0}$ a segmentation mask at t_0 for the patient e obtained by the proposed method. Then, the area criterion is defined as follows:

$$A^e_{t_{k/0}} = \frac{\text{card}\{x : \ B^e_{t_{k/0}}(x) = 1 \ \wedge \ C^e_{t_0}(x) = 1\}}{\text{card}\{x : \ C^e_{t_0}(x) = 1\}}. \tag{9}$$

2.4.2 multi-level change detection: homogeneity

To measure the homogeneity we propose to use a multi-level change detection approach. We consider the statistical framework based on a contextual hypothesis testing similarly to the "Statistical Parametric Mapping (SPM)" ([17, 44]) approach in functional MRI. We model the change map as a realization of a Gaussian random field under the null hypothesis. Then a multi-scale analysis partition this map into regions and a probability is assigned to each region. These probabilities measure the likelihood for the regions under the null hypothesis to be a realization of the random field. For our application, we apply this methodology on each change map $F^e_{t_{k/0}}$. To model $F^e_{t_{k/0}}$ as a Gaussian random field, we need to normalize it. Then we apply the multi-scale analysis to get regions and calculate probabilities.

Map normalization The SPM methodology, makes the hypothesis that data follow a highly correlated standard Gaussian distribution. To satisfy this assumption, we smooth the image $F^e_{t_{k/0}}$ with a Gaussian convolution filter to increase its spatial correlation. Second, as the interest area of $F^e_{t_{k/0}}$ comes from a unique class, its histogram can be modelled as monomodal in the hypothesis of no changes. We use a histogram specification to transform it into a Gaussian distribution. We center and reduce the specified histogram to get a standard Gaussian distribution. In the following we will denote $F^{Ne}_{t_{k/0}}$ the normalized change map.

Change map thresholding On the normalized map $F^{Ne}_{t_{k/0}}$ we apply the Gaussian field theory ([2]). An interesting characteristic of a Gaussian field F of size S is the DT (Differential Topology) characteristic. We first define the excursion set A_u for each threshold u as follows:

$$A_u = \{s \in S, F(s) \geq u\}. \tag{10}$$

For each connected component $C_u(j)$ of A_u, the DT characteristic is given by:

$$\chi(C_u(j)) = (-1)^{(d-1)} \sum_{k=0}^{d-1} (-1)^k \chi_k(C_u(j)) \tag{11}$$

where u is a given threshold, d the dimension of the field and $\chi_k(A_u)$ the number of points $s \in C_u(j)$ satisfying the conditions:

- $F(s) = u$
- $F^{(i)}(s) = 0, \forall i \in [1, d-1]$
- $F^{(d)}(s) > 0$
- the $(d-1) \times (d-1)$ matrix of second order partial derivatives of F(s) has exactly k negative eigenvalues.

In the above expression, $F^{(i)}$ denotes the i^{th} partial derivative of F with respect to variable s. In the 2D case (i.e. $d = 2$), for a given threshold u the expectation of the DT characteristic of a centred and reduced Gaussian field can be written as follows ([2]):

$$\mathbb{E}_{\chi(A_u)} = \mathcal{S}(2\pi)^{\frac{3}{2}} |\Lambda|^{\frac{1}{2}} \sigma^{-3} u e^{-\frac{u^2}{2\sigma^2}} \tag{12}$$

with $\mathcal{S}$ the surface of the field, Λ the 2×2 correlation matrix of the field derivatives, and $\sigma^2 = \mathbb{E}(F^2)$.

In order to estimate the likelihood of a connected component of $F^{Ne}_{t_{k/0}}$ above the threshold u under the Gaussian field assumption, we compare its characteristic with the DT characteristic. Two parameters can be used to compute this likelihood: the maximum intensity and the spatial extent of the considered neighbourhood above u.

Let x_0 denotes the maximum intensity of $R^{x_0}_u$ a region above the threshold u. The likelihood of this region under the Gaussian field assumption is ([44]):

$$P(R^{x_0}_u \in F) = \frac{\mathbb{E}_{\chi(A_{x_0})}}{\mathbb{E}_{\chi(A_u)}} = \frac{x_0}{u} e^{\frac{u^2 - x_0^2}{2}}. \tag{13}$$

Let S_0 denotes the spatial extent of $R^{S_0}_u$ a region above the threshold u. The expectation for this region under the Gaussian field assumption is ([44]):

$$\mathbb{E}(R^{S_0}_u \in F) = \frac{\mathbb{E}(N_u)}{\mathbb{E}_{\chi(A_u)}}, \tag{14}$$

where $\mathbb{E}(N_u)$ is the expectation to have N_u pixels above the threshold u in a Gaussian random field F. As F follows a standard normal distribution,

$$\mathbb{E}(N_u) = \mathcal{S} \int_u^\infty (2\pi)^{\frac{1}{2}} e^{-\frac{x^2}{2}} = \mathcal{S}\Phi(-u) \tag{15}$$

then,

$$\mathbb{E}(R^{S_0}_u \in F) = \frac{\Phi(-u)}{(2\pi)^{\frac{3}{2}} |\Lambda|^{\frac{1}{2}} u e^{-\frac{u^2}{2}}}. \tag{16}$$

Algorithm 1 Multi-scale analysis

Require: I_0, I_k, M_{p_0}
$F^e_{t_{k/0}} = I_k - I_0$
Smooth and normalize $F^e_{t_{k/0}}$ to get $F^{Ne}_{t_{k/0}}$ (see section 2.4.2)
Choose a family $U = u_1, ..., u_{n_u}$
for $i = 1$ to n_u **do**
  Compute the connected components above u_i
  for each connected component **do**
    Calculate x_0 and S_0
    Calculate $P(R_u^{x_0,S_0} \in F)$ with eq. 18
  end for
end for
Merge the obtained connected clusters obtained with the family U to get $SM^e_{t_k}$

Thus, the number of pixels in a region above the threshold u has an exponential distribution with the parameter $\lambda_u = 1/\mathbb{E}(R_u^{S_0} \in F)$. We can then write the probability of a region $R_u^{S_0}$ to appear in a realization of the Gaussian field by:

$$P(R_u^{S_0} \in F) = e^{-\lambda_u S_0} = e^{\frac{(2\pi)^{\frac{3}{2}} |\Lambda|^{\frac{1}{2}} S_0 u e^{-\frac{u^2}{2}}}{\Phi(-u)}} \tag{17}$$

Finally, the probability for a spatial region above u with maximum intensity x_0 and spatial extend S_0, to be in F is:

$$P(R_u^{x_0,S_0} \in F) = min(P(R_u^{x_0} \in F), P(R_u^{S_0} \in F)). \tag{18}$$

Thereby, to detect changes between t_0 and t_k we define a family of n thresholds $U = \{u(1), ..., u(n)\}$. For each $u(i)$, we consider the connected components $R_{u(i)}$ of the map $F^e_{t_{k/0}}$ thresholded at $u(i)$. Then we assign a probability $P(R_{u(i)}^{x_0,S_0} \in F)$ to each region $R_{u(i)}^{x_0,S_0}$ (see Figure 5). This process gives a statistical change map between t_0 and t_k for each threshold of U. A final change map denoted $SM^e_{t_k}$ between t_0 and t_k is obtained by merging the change maps obtained for each thresholds of U. This merging consists of stacking the changes maps sorted in increasing value of u. This stacking method allows to get all the obtained regions in a single map showing level set of the pathology changes. The algorithm 1 summarizes the procedure for obtaining SM_{t_k} between I_0 and I_k for a given patient.

normalization The statistical approach proposed in Section 2.4.2 makes the hypothesis that changes are rare events. In our images, a large part of the pathological area can change. That breaks this assumption (see Figure 6 d to f). This result is an underdetection of the changes. Homogeneity variations cannot be quantified in these maps. We propose then to introduce the following normalization:

$$F^{Nde}_{t_{k/0}} = F^{Ne}_{t_{k/0}} + \frac{1}{N_r} \sum_{r=t_0}^{t_k} \left(\mu_{C_{t_0}}(F^{Ne}_{t_{r/0}}) - \mu_{C_{t_0}}(F^{Ne}_{t_{1/0}}) \right) \tag{19}$$

where N_r is the number of time measurements between t_0 and t_k. $\mu_{C_{t_0}}(F^{Ne}_{t_{r/0}})$ (respectively $\mu_{C_{t_0}}(F^{Ne}_{t_{1/0}})$) denotes the average of the pixel intensities in the ROI of $F^{Ne}_{t_{r/0}}$ (respectively $F^{Ne}_{t_{1/0}}$) delimited by C_{t_0}. As the normalized data using Equation (19) do not respect the rare event

Figure 5: Muli-scale analysis illustration. Green: map thresholded at $u(1)$, red: map thresholded at $u(2) > u(1)$, and gray: regions below $u(1)$ and $u(2)$. A probability is assigned to each region above the thresholds u_1 and u_2.

hypothesis, the computed statistics are biased. Nevertheless, we do not use the absolute values of these statistics but their spatial repartition. Figure 6 g to i shows $SM^e_{t_k}$ cartographies obtained with the normalization given by Equation (19). The pathology evolution is now visible during the time sequence.

Homogeneity criterion To quantify the homogeneity evolution we integrate the multi-scale change maps into a scalar value. To do so, we define a function $f^e_s(p)$ that represents the spatial repartition of the changes by:

$$f^e_s(p) = \frac{card\{x : SM^e_{t_k}(x) \leq p\}}{card\{SM^e_{t_k}\}}. \tag{20}$$

$f^e_s(p)$ is an increasing function from $[0,1]$ to $[0,1]$. To be able to represent the homogeneity with respect to the surface, we prefer to work with $f^{pe}_{t_k}(s)$ that is the inverse function of $f^e_s(p)$. A linear interpolation is used to calculate $f^{pe}_{t_k}(s)$ with a regular sampling along (s_i). $f^{pe}_{t_k}(s)$ is also an increasing function from $[0,1]$ to $[0,1]$. In the following, we use $f^{pe}_{t_k}(s)$ for both calculating a graphical representation of the homogeneity criterion (see section 5), and a scalar $H^e_{t_k}$ describing the homogeneity of the patient e at time t_k.

To integrate $f^{pe}_{t_k}(s)$ in a scalar $H^e_{t_k} \in [0,1]$, we compute:

$$H^e_{t_k} = 2\int_0^1 |f^{pe}_{t_k}(s) - s| ds \simeq \sum_i 2|f^{pe}_{t_k}(s_i) - s_i|(s_{i+1} - s_i). \tag{21}$$

$H^e_{t_k}$ represents the area between the function $f^{pe}_{t_k}(s)$ and the identity function $Id(s) = s$. Id corresponds to the most heterogeneous repartition that can be expected. Then, more homogeneous changes yield higher values $H^e_{t_k}$.

3 Experimental settings

In this section, we introduce the database constituted of two clinical studies that is used for validating the proposed method and we expose the protocol that we use to evaluate it.

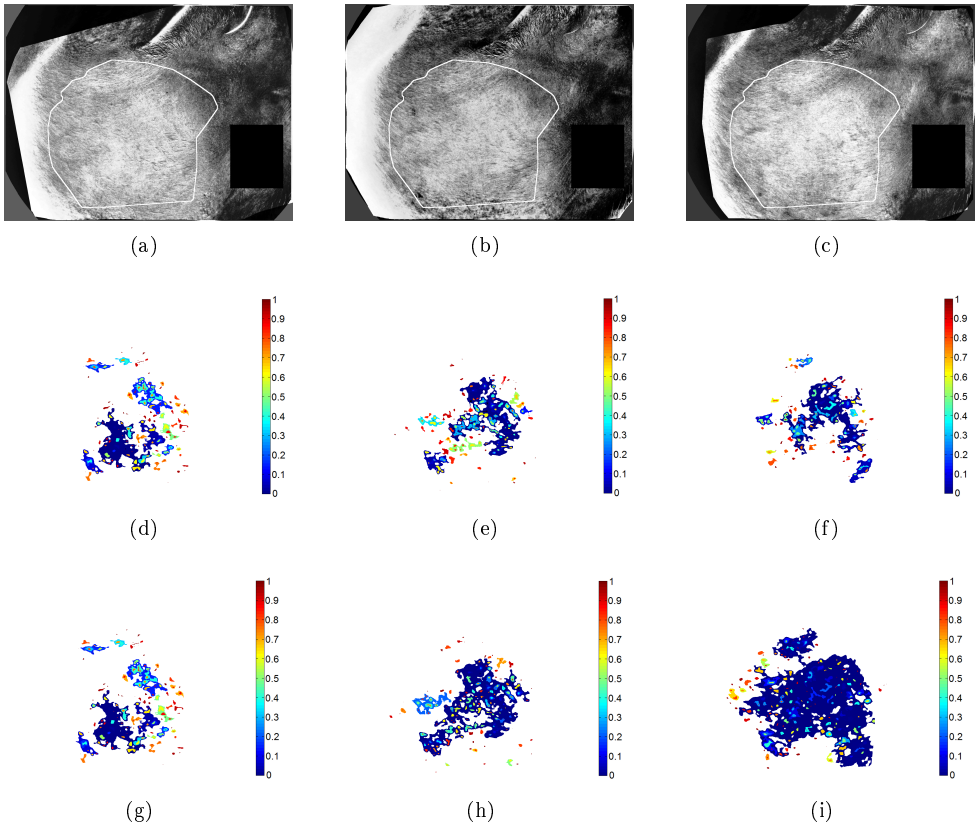


Figure 6: Changes map of the differential score for a patient whose pathology decreases during the treatment period. a-c) equalized difference image on the spectral feature map between each time t_1, t_2 and t_3 and t_0, d-f) homogeneity maps $SM^e_{t_k}$ for the three time t_1, t_2 and t_3 without using the normalization of Equation (19) w.r.t g-i) homogeneity maps $SM^e_{t_k}$ for the three time t_1, t_2 et t_3 using the normalization of Equation (19).

3.1 Database

The database we use it made of 2 clinical studies. The first one involves three active treatments that are denoted S_t, A_{d1} and A_{d2}. Each treatment is tested on a population of 16 patients. For each patient, one cheek receives one of the treatments S_t, A_{d1} or A_{d2} and the second cheek receives a comparator. The study lasts 3 months and two multispectral images are taken for each patient (one per cheek) every month. Thus, we have $3 * 16 * 2 = 96$ series of four images corresponding to times t_0 = baseline, t_1 = after one month, t_2 = after two months and t_3 = after three months. The total amount of images of the first clinical study is therefore equal to 384. The images are of a size of 1200*960 pixels, and contain 18 spectral bands with wavelengths varying from 405 nm to 970 nm. For this first clinical study, the dermatologists conclusions state that S_t has no efficacy whereas A_{d1} and A_{d2} have efficacy. Furthermore, A_{d2} has a stronger effect than A_{d1}.

The second clinical study compares two treatments denoted T and A. Each treatment is tested on a population of 22 patients. Thus, the clinical study involve 44 patients. For each patient, one cheek receives one of the treatments T or A and the second cheek receives a comparator. The protocol of this study is the same as for the first clinical study. Then, this study involve 352 multispectral images.

3.2 Evaluation protocol

The proposed "differential score" composed of three criteria (darkness, area and homogeneity) appears as a numeric equivalent of the MASI criterion defined by dermatologists. We evaluate the three components individually. In order to evaluate the accuracy of the severity changes of this new score, we used it on the clinical studies cited above. For each of these two clinical studies, dermatologists scored the MASI for each patient at each visit. From the obtained table of scores, we are able to compute statistical hypothesis tests to evaluate if a treatment had an effect or not, due to the MASI criterion. After the hypothesis testing, if the p-value of the test is below 0.05, we consider that the score significantly changed between two measurements and then conclude that the treatment had an effect. In the experiment section, we run the proposed "differential score" on the whole dataset of the clinical studies. With the obtained table of scores, we can perform the same statistical testing as with the dermatologists scores. If the results of the statistical tests using the "differential score" and the dermatologists diagnosis match, then we conclude that the proposed "differential score" is accurate. As they are parameters to calibrate in order to obtain the "differential score", we propose to use the two clinical studies as follows: the first clinical study is used to tune the parameters. This means that, we experimentally test different sets of parameters to find which one gives conclusions similar to the dermatologists scoring. Then we run the "differential score" on the second clinical study with the same set of parameters, to validate the proposed method. If the same set of parameters provides the same conclusions as the dermatologists on the second clinical study, we conclude that the proposed "differential score" accurately quantify the Melasma evolution.

4 Results

4.1 Darkness

In order to compare the features extraction methods detailed in section 2.3 we calculate the spectral feature for each treatment S_t, A_{d1} and A_{d2} of the first clinical study individually. Then we compute the features in all the images of the study on the four segmented regions $I_{t_k}^{pS}$, $I_{t_k}^{hS}$,

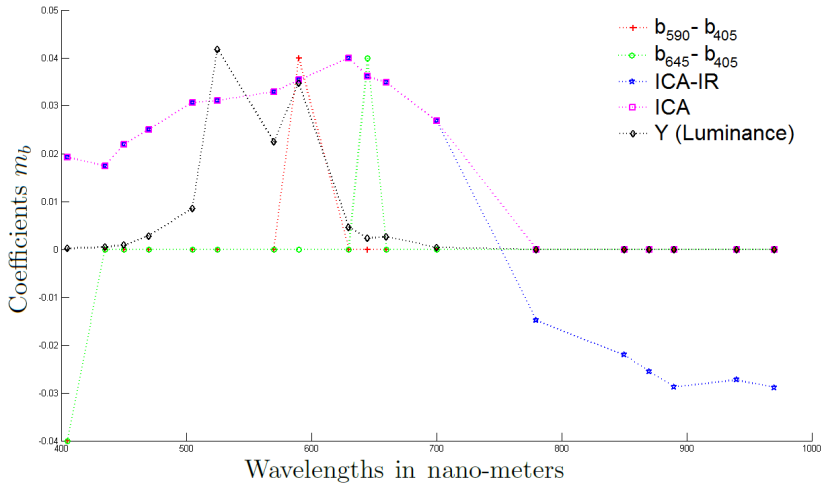


Figure 7: Spectral features obtained by the studied methods.

$I_{t_k}^{pC}$, $I_{t_k}^{hC}$. this results in a darkness measurement $D_{t_k}^{e}$ for each patient. We finally apply a Wilcoxon hypothesis test for each treatment between t_0 and t_k to get a quantification of the treatment efficacy. Table 1 shows the p-values obtained by the Wilcoxon test with the proposed method, with the ICA and with the luminance that we use as a reference feature. Equation (5) is used for normalization. As one can see, all the features agree with the clinical conclusions. The difference is in how fast we can detect an efficacy of the treatment. The method based on the objective function optimization gives a faster detection compared to the ICA-based method and the luminance. To understand this difference, we plot the features obtained by all the methods in Figure 7. As the luminance coefficient is not linear, we use its linear approximation Y. We observe that the feature obtained by f selects a specific band at 590 nm for the treatments S_t, and A_{d2} and the band 645 nm for the treatment A_{d1}. ICA selects spectral bands mainly around 600 nm, but sets significant coefficients to wavelengths between 400 and 600 nm. The Y feature selects bands in the 500-600 nm wavelengths. Due to the Stamatas et al. model for pigmentation ([57]) we should focus on the wavelengths between 600 and 700 nm. Thus, a possible interpretation to explain why the ICA based method and the Y features detect later the treatments efficacy than the feature estimated by f, is there high coefficients in the spectral ranges out of 600-700 nm. Table 2 shows the results obtained with the Wilcoxon test on the same studies with the normalization given by Equation (4). The results are comparable to the ones obtained with the normalization given by Equation (5). This means that the calibration of the camera is robust and that the normalization using the healthy area is not mandatory. As this normalization can introduce noise due to the variation of the healthy area, we prefer to use the normalization given by Equation (4). To validate the spectral features calculated on the first clinical study, we apply them on the second study described in section 3.1. The p-values of the Wilcoxon tests on the second study are shown in Table 3. For f we choose the feature $b_{590} - b_{405}$. Obtained results are totally coherent with the clinical analysis. We can then conclude that the proposed darkness criterion using the f feature is accurate and consistent with the dermatologist diagnosis.

Table 1: P-values obtained by the Wilcoxon test on the first clinical study with the normalization of Equation (5) for the darkness criterion.

Feature	Treatment	$t_1 - t_0$	$t_2 - t_0$	$t_3 - t_0$
L	S_t	$7.959\ 10^{-1}$	$5.014\ 10^{-1}$	$7.173\ 10^{-1}$
	A_{d1}	$3.793\ 10^{-1}$	$1.128\ 10^{-2}$	$8.793\ 10^{-2}$
	A_{d2}	$9.798\ 10^{-2}$	$3.204\ 10^{-3}$	$5.312\ 10^{-4}$
f	S_t	$1.476\ 10^{-1}$	$1.089\ 10^{-1}$	$7.563\ 10^{-1}$
	A_{d1}	$5.571\ 10^{-2}$	$3.400\ 10^{-2}$	$4.373\ 10^{-2}$
	A_{d2}	$8.360\ 10^{-3}$	$3.204\ 10^{-3}$	$9.350\ 10^{-4}$
ICA	S_t	$5.349\ 10^{-1}$	$8.767\ 10^{-1}$	$9.176\ 10^{-1}$
	A_{d1}	$4.080\ 10^{-1}$	$2.987\ 10^{-2}$	$9.798\ 10^{-2}$
	A_{d2}	$7.873\ 10^{-2}$	$5.233\ 10^{-3}$	$9.350\ 10^{-4}$

Table 2: P-values obtained by the Wilcoxon test on the first clinical study with the normalization of Equation (4) for the darkness criterion.

Feature	Treatment	$t_1 - t_0$	$t_2 - t_0$	$t_3 - t_0$
L	S_t	$7.959\ 10^{-1}$	$6.791\ 10^{-1}$	$6.416\ 10^{-1}$
	A_{d1}	$1.306\ 10^{-2}$	$2.289\ 10^{-2}$	$9.798\ 10^{-2}$
	A_{d2}	$1.788\ 10^{-1}$	$7.169\ 10^{-3}$	$1.918\ 10^{-3}$
f	S_t	$7.173\ 10^{-1}$	$5.349\ 10^{-1}$	$5.014\ 10^{-1}$
	A_{d2}	$2.289\ 10^{-2}$	$1.737\ 10^{-2}$	$3.400\ 10^{-2}$
	A_{d3}	$1.997\ 10^{-2}$	$1.609\ 10^{-3}$	$9.725\ 10^{-3}$
ICA	S_t	$6.416\ 10^{-1}$	$5.694\ 10^{-1}$	$5.014\ 10^{-1}$
	A_{d1}	$1.997\ 10^{-2}$	$3.400\ 10^{-2}$	$8.793\ 10^{-2}$
	A_{d2}	$1.476\ 10^{-1}$	$7.169\ 10^{-3}$	$1.123\ 10^{-3}$

Table 3: P-values obtained by the Wilcoxon test on the second clinical study for the darkness criterion.

Test	Treatment	$t_1 - t_0$	$t_2 - t_0$	$t_3 - t_0$
L	A	$8.455\ 10^{-1}$	$7.210\ 10^{-1}$	$1.270\ 10^{-1}$
	T	$9.741\ 10^{-1}$	$9.397\ 10^{-3}$	$1.637\ 10^{-3}$
f	A	$1.116\ 10^{-1}$	$5.373\ 10^{-1}$	$9.777\ 10^{-2}$
	T	$4.750\ 10^{-1}$	$3.859\ 10^{-3}$	$9.337\ 10^{-3}$
ICA	A	$7.210\ 10^{-1}$	$5.589\ 10^{-1}$	$1.627\ 10^{-1}$
	T	$8.966\ 10^{-1}$	$6.389\ 10^{-3}$	$3.478\ 10^{-3}$

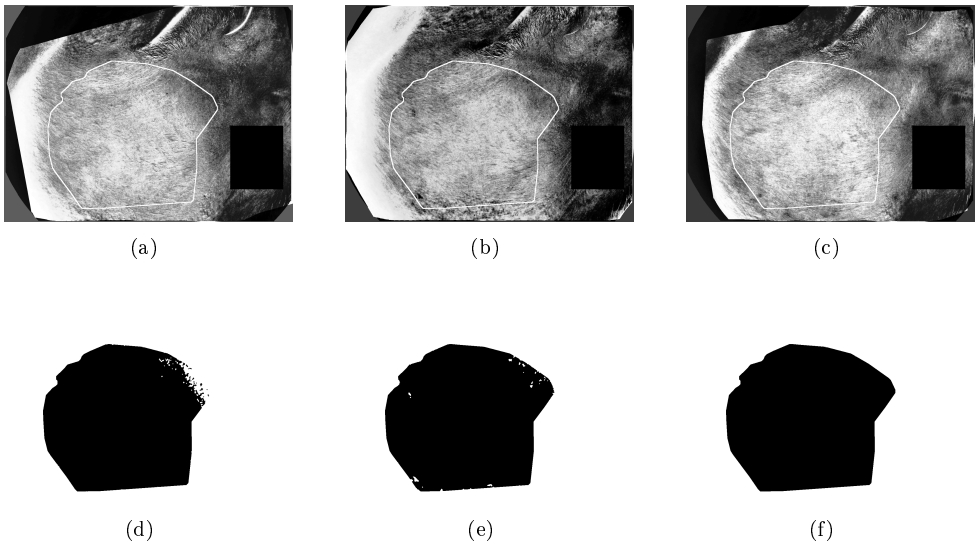


Figure 8: Changes map for a patient whose pathology decreases during the treatment period. a-c) equalized difference image on the spectral feature map between each time t_1, t_2 and t_3 and t_0, d-f) change detection map used to measure the area for each times t_1, t_2 and t_3. Changed areas are drawn in black.

Table 4: p-values of the Wilcoxon tests on the second clinical study for the area criterion.

Criterion	Treatment	$t_1 - t_0$	$t_2 - t_0$	$t_3 - t_0$
Area	A	$2.840\ 10^{-1}$	$3.986\ 10^{-1}$	$1.045\ 10^{-1}$
	T	$4.552\ 10^{-1}$	$3.478\ 10^{-3}$	$2.438\ 10^{-4}$

4.2 Area

The area criterion does not have parameters to calibrate. Thus, we can test this criterion directly on the second clinical study. Figure 8 d-f) shows obtained results for an image time series. For this patient, dermatologists clinically measure a decrease of the pathology. As one can see, the area measured by the algorithm shows a significant evolution at time t_1 and then a gradual evolution of the pathological area. To validate this area measurement, we calculate change areas $A^e_{t_{k/0}}$ for all the patients of the second clinical studies. As for the darkness measurement, we compute the Wilcoxon test for each treatment between t_0 and t_k in the area measurements. Results are shown in Table 4. Results show that the treatment T has an effect whereas the treatment A does not has any. This is in agreement with the dermatologist diagnosis.

4.3 Homogeneity

To validate the proposed homogeneity criterion, we first calibrate the parameters of the model using images series from patients of the first clinical study. These patients were selected to be identified to have a decreasing pathology by the dermatologists. We then apply the homogeneity calculation on the second study. The homogeneity measure makes sense only when some changes appear in the image between t_0 and t_k. A margin is then fixed at 10% of the pathological

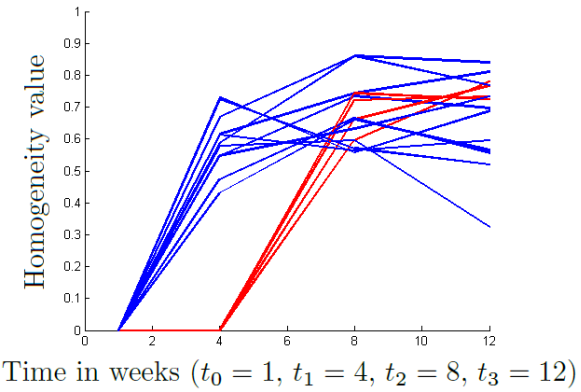


Figure 9: Time sequences of the homogeneity measurements for the patients of the group for which an evolution has been measured by the area and darkness criteria.

area changes below which we do not consider the criterion $H^e_{t_k}$. To qualitatively evaluate the homogeneity criterion, we plot its evolution for the population of patients whose pathology evolves. Due to the darkness and the area criteria, the considered patients are those who received the treatment T in the second study. Figure 9 shows time plots of the homogeneity measurement for each patient of the group receiving the treatment T. If the treatment has an effect, we expect to see the homogeneity measurement growing. In this population of 22 patients, the H criterion is not interpretable for 7 patients. Indeed, for five patients less than 10% of the pathological area changes at time $t_2 = 8$ or $t_3 = 12$ and for two patients there were less than 10% of the pathological area that changes in all the time sequence. For the other fifteen patients we globally observe the expected growth of the homogeneity with time. For most patients (in blue) main changes occurred in the second measurement time $t_1 = 4$, and for 4 patients (in red), main changes occurred in time $t_2 = 8$. These results show that the homogeneity criterion gives relevant interpretation of disease regression during treatment. This criterion is by definition interpretable only on patients whose pathology has evolved.

5 Discussions and Conclusions

5.1 Discussions

In this report we proposed a "differential score" that allows computing the evolution of the Melasma severity along a clinical study. This "differential score" is composed of three independent scores that are the darkness, the area and the homogeneity. The MASI described in [41] and [26] calculates a linear combination of darkness, area and homogeneity measurements (see Equations (1) and (2)). For the proposed "differential score", such a combination does not have any meaning. Indeed, the three criteria have different dynamics, and combining them can bias the final measurement. That is why we prefer to use the three measurements independently. Besides, in practice the homogeneity of the MASI is usually not used by dermatologists and removed from the global score.

Furthermore, the proposed "differential score" summarizes the variation of pathology between two images with 3 numbers that represent different aspects of the treatment efficacy. To make the "differential score" easily readable for an individual patient we use a graphical representation as follows:

(a) (b) (c)

Figure 10: Darkness graphical representation. a) for a patient whose pathology became darker, b) for a patient with stationary darkness, c) for a patient whose pathology decreases.

(a) (b) (c)

Figure 11: Area graphical representation. a) for a patient with stationary pathological area, b) for a patient whose pathological area decreases by 26%, c) for a patient whose pathological area decreases by 86%.

Darkness representation The proposed darkness criterion is a differential measure between two monochrome images. Its theoretical magnitude is then $[-255, 255]$ for 8 bit images. In practice, the magnitude observed in our data set is in the interval $[-12, 12]$. To visualize the changes of D, we compute D_G the darkness value used for the graphical representation by shrinking the representation area using a curve with a shape given by:

$$D_G = \frac{D}{\sqrt{1 + D^2}}. \tag{22}$$

The darkness bar contains a chip with a color from red (-12) to blue (+12). This chip position on the bar depends on the D_G value from -12 to 12. Figure 10 shows three examples of darkness representation bars.

Area representation The area criterion we proposed has values in the interval $[0, 1]$. The graphical representation we have chosen is then a percentage bar. Figure 11 shows examples of graphical representation of the area criterion.

Homogeneity representation For the graphical representation of the homogeneity criterion, we propose to use $f^{pe}_{t_k}(s)$. In fact this information is richer than $H^e_{t_k}$. The graphical bar of homogeneity is then a plot of $f^{pe}_{t_k}(s)$ with a gradient color. Values of $f^{pe}_{t_k}(s)$ close to zero are shown in blue and values of $f^{pe}_{t_k}(s)$ close to one are given in red. Intermediate values are represented with the "jet" colormap. If the gradient is all red, that means there is no homogeneity variation. The slower the gradient changes from blue to red, the more heterogeneous the variation is. Figure 12 shows examples of the homogeneity bar.

Differential score representation Finally, a differential score representation can be the combination of the three bars of darkness, area and homogeneity. An example for a patient whose

(a) (b) (c)

Figure 12: Homogeneity graphical representation. a) for a patient without evolution of the pathology, b) for a patient with heterogeneous variation of the pathology, c) for a patient with homogeneous variation of the pathology.

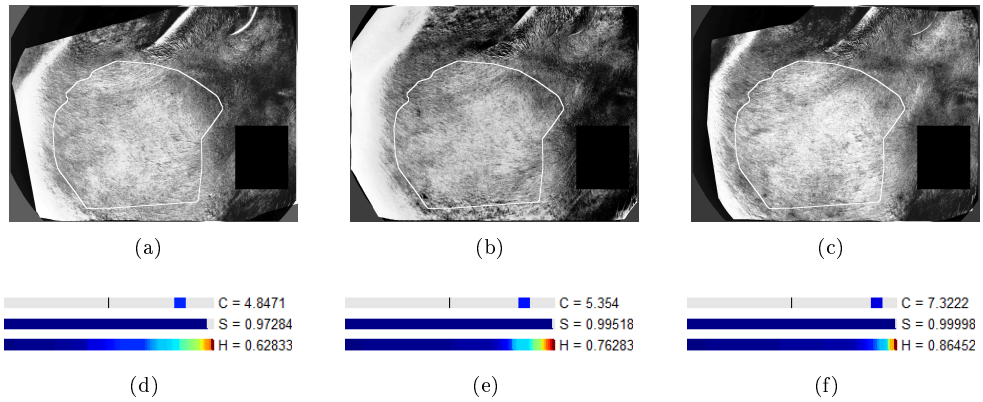


Figure 13: Graphical representation of the differential score for a patient whose pathology decreases during the treatment period. a-c) equalized difference image on the spectral feature map: $t_1 - t_0$, $t_2 - t_0$ and $t_3 - t_0$, d-f) graphical representation of the differential score for the three times t_1, t_2 and t_3.

pathology decreases is shown in Figure 13 and an example for a patient whose pathology does not decrease is shown in Figure 14. As one can see, with this graphical representation, we can both quantify and qualify the patient pathology evolution.

5.2 Conclusion

In this report, we proposed a strategy to calculate the severity of skin hyperpigmentation using multispectral images. The proposed criterion called "differential score" has been developed to be a computer score related to the clinical MASI criterion. The proposed differential score has been tested on a full clinical study and has given relevant results with respect to the clinical analysis. Advantages of our method are that it can estimate the spectral feature of diseases through the training phase using the reference data set, and that no prior knowledge is required in the spatial analysis. As such, our method can be adapted to analyze other pathologies such as vitiligo, scars or rosacea.

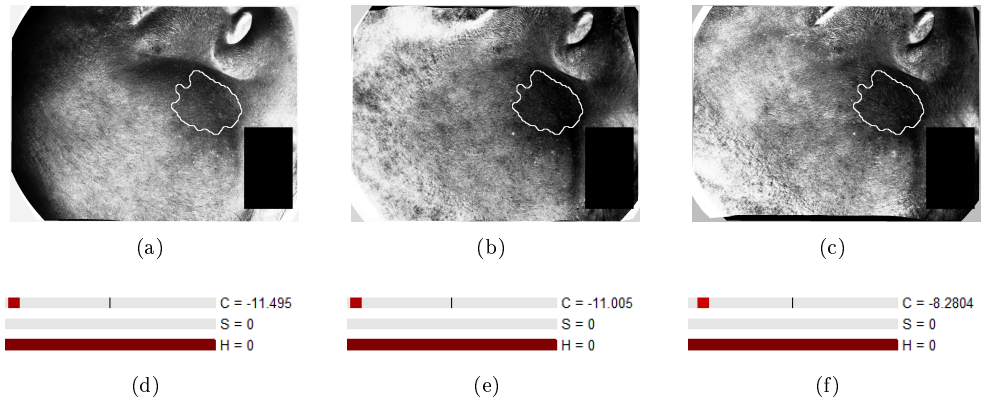


Figure 14: Graphical representation of the differential score for a patient whose pathology did not decrease during the treatment period. a-c) equalized difference image on the spectral feature map: $t_1 - t_0$, $t_2 - t_0$ and $t_3 - t_0$, d-f) graphical representation of the differential score for the three times t_1, t_2 et t_3.

References

[1] T. Aach and A. Kaup. Statistical model-based change detection in moving video. *Signal Process.*, 31:165–180, Mar. 1993.

[2] R.J. Alder. *The Geometry of Random Fields*. Wiley, 1981.

[3] R. Balkrishnan, A.J. Mcmichael, F.T. Camacho, F. Saltzberg, T.S. Housman, S. Grummer, S.R. Feldman, and M.M. Chren. Development and validation of a health-related quality of life instrument for women with melasma. *British Journal of Dermatology*, 149:572–577, 2003.

[4] J.E. Ball, L.M. Bruce, and N.H. Younan. Hyperspectral pixel unmixing via spectral band selection and DC-insensitive singular value decomposition. *Geoscience and Remote Sensing Letters, IEEE*, 4, Juillet 2007.

[5] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. on Geoscience and Remote Sensing*, 40:480–491, 2005.

[6] J. Bezdek. *Pattern recognition with fuzzy objective functions*. Plenum Press, New York., 1981.

[7] J. W. Boardman, F.A. Kruse, and R.O. Green. Mapping target signatures via partial unmixing of AVIRIS data. In *in Summaries of JPL Airborne Earth Science Workshop*, 1995.

[8] L. Bruzzone and D. F. Prieto. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans. Image Processing*, 11:452–466, Avr. 2002.

[9] J.F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11:157–192, 1999.

[10] G. Celeux and G. Govaert. A classification EM algorithm for clustering and two stochastic versions. *Comput. Stat. Data Anal.*, 14:315–332, Oct. 1992.

[11] C. Chang, C.C. Wu, and C.T. Tsai. Random N-Finder (N-FINDR) endmember extraction algorithms for hyperspectral imagery. *IEEE Trans. on Image Processing*, 20(3):641–656, March 2011.

[12] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[13] A. Cornuéjols and L. Miclet. *L'apprentissage artificiel concepts et algorithmes 2ème édition.* Eyrolles, 2009.

[14] I.S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with Bregman divergences. In *Neural Information Proc. Systems*, 2005.

[15] N. Dobigeon, S. Moussaoui, M. Coulon, J.Y. Tourneret, and A.O. Hero. Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Trans. Signal Processing*, 57(11):4355–4368, Nov. 2009.

[16] M. Fauvel. Spectral and spatial methods for the classification of urban remote sensing data. In *PhD Thesis.* Institut National Polytechnique de Grenoble. France, 2007.

[17] K.J. Friston, J. Ashburner, S.J. Kiebel, T.E. Nichols, and W.D. Penny, editors. *Statistical Parametric Mapping: The Analysis of Functional Brain Images.* Academic Press, 2007.

[18] A.A. Green, M. Berman, P. Switzer, and M.D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosciences Remote Sensing*, 26, 1988.

[19] G. Hazel. Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. on Geoscience and Remote Sensing*, 38:1199–1211, 2000.

[20] G. F. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. on Information Theory*, 14:55–63, 1968.

[21] M.E. Hurley, I.L. Guevara, R.M. Gonzales, and A.G. Pandya. Efficacy of glycolic acid peels in the treatment of melasma. *Archives of Dermatology*, 138:1578–1582, 2002.

[22] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10:626–634, 1999.

[23] J. Inglada and G. Mercier. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Trans. on Geoscience and Remote Sensing*, 45:1432–1445, mai 2007.

[24] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31:264–323, Sept. 1999.

[25] L.O. Jimenez, J.L. Rivera-Medina, E. Rodriguez-Diaz, E. Arzuaga-Cruz, and M. Ramirez-Velez. Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Trans. on Geoscience and Remote Sensing*, 43:844–851, avril 2005.

[26] CK Kimbrough-Green, CE Griffiths, LJ Finkel, TA Hamilton, SM Bulengo-Ransby, CN Ellis, and et al. Topical retinoic acid (tretinoin) for melasma in black patients. a vehicle-controlled clinical trial. *Arch Dermatol*, 130:727–33, 1994.

[27] D.A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, New Jersey,, 2003.

[28] C. Lee and D.A. Landgrebe. Feature extraction based on decision boundaries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:388–400, Apr 1993.

[29] C. Lee and D.A. Landgrebe. Decision boundary feature extraction for neural networks. *IEEE Trans. on Neural Networks*, 8:75–83, Jan 1997.

[30] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[31] L. Li and M. K. H. Leung. Integrating intensity and texture differences for robust change detection. *IEEE Trans. Image Process.*, 11:105–112, Fev. 2002.

[32] S. Liu, C. Fu, and S. Chang. Statistical change detection with moments under time-varying illumination. *IEEE Trans. Image Process.*, 7:1258–1268, Sep. 1998.

[33] D. Lu, P. Mausel, E. Brondízio, and E. Moran. Change detection techniques. *International Journal of Remote Sensing*, 25(12):2365–2401, 2004.

[34] R.M.P. Manaloto and T. Alster. Erbium:YAG laser resurfacing for refractory melasma. *Dermatologic Surgery*, 25:121–123, 1999.

[35] P. Mather. *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley and sons, 2004.

[36] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the theory of statistics*. McGraw-Hill, 1974.

[37] A. Myronenko and X. Song. Intensity-based image registration by minimizing residual complexity. *IEEE Trans. on Medical Imaging*, 29(11):1882–1891, 2010.

[38] J. M. P. Nascimento and J. M. Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 43:898–910, Apr. 2005.

[39] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.

[40] A. Pandya, M. Berneburg, J. Ortonne, and M. Picardo. Guidelines for clinical trials in melasma. pigmentation disorders academy. *Br J Dermatol*, 156 Suppl 1, 2006, 2006.

[41] A.G. Pandya, L.S. Hynan, R. Bhore, F.C. Riley, I.L. Guevara, P. Grimes, J.J. Nordlund, M. Rendon, S. Taylor, R.W. Gottschalk, N.G. Agim, and J.P. Ortonne. Reliability assessment and validation of the melasma area and severity index (MASI) and a new modified MASI scoring method. *Journal of the American Academy of Dermatology*, 64(1):78 – 83.e2, 2011.

[42] A. Plaza and C.I. Chang. A Fast Iterative Algorithm for Implementation of Pixel Purity Index. *IEEE Geoscience and Remote Sensing Letters*, 3(1):63–67, 2006.

[43] A. Plaza, P. Martinez, R. Perez, and J. Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Trans. on Geoscience and Remote Sensing*, 40:2025–2041, 2002.

[44] J.B. Poline, K.J. Worsley, A.C. Evans, and K.J. Friston. Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage*, 5(2):83–96, 1997.

[45] H. V. Poor. *An Introduction to Signal Detection and Estimation, 2nd ed.* New York: Springer-Verlag, 1994.

[46] S. Prigent, X. Descombes, D. Zugaj, P. Martel, and J. Zerubia. Multi-spectral image analysis for skin pigmentation classification. In *Proc. IEEE International Conference on Image Processing (ICIP)*, Hong-Kong, China, September 2010.

[47] S. Prigent, D. Zugaj, X. Descombes, P. Martel, and J. Zerubia. Estimation of an optimal spectral band combination to evaluate skin disease treatment efficiency using multi-spectral images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, September 2011.

[48] Sylvain Prigent, Xavier Descombes, Didier Zugaj, Laurent Petit, Anne-Sophie Dugaret, Philippe Martel, and Josiane Zerubia. Classification of skin hyper-pigmentation lesions with multi-spectral images. Research Report RR-8105, INRIA, October 2012.

[49] Sylvain Prigent, Xavier Descombes, Didier Zugaj, Laurent Petit, Anne-Sophie Dugaret, Philippe Martel, and Josiane Zerubia. Skin lesion evaluation from multispectral images. Research Report RR-8136, November 2012.

[50] R.J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. on Image Processing*, 14(3):294 – 307, 2005.

[51] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia. Texture feature analysis using a Gauss-Markov model in hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing*, 42:1543–1551, 2004.

[52] M. Riedmann and E.J. Milton. Supervised band selection for optimal use of data from airborne hyperspectral sensors. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 3, 2003.

[53] E. Rignot and J. Van Zyl. Change detection techniques for ERS-1 SAR data. *IEEE Trans. Geosci. Remote Sens.*, 31:896–906, Juil. 1993.

[54] K. Skifstad and R. Jain. Illumination independent change detection for real world image sequences. *Comput. Vis. Graph. Image Process.*, 46:387–399, 1989.

[55] Pierre Soille. *Morphological Image Analysis: Principles and Applications.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 edition, 2003.

[56] G. N. Stamatas, B. Z. Zmudzka, N. Kollias, and J. Z. Beer. Non-invasive measurements of skin pigmentation in situ. *Pigment cell res*, 17:618–626, 2004.

[57] G. N. Stamatas, B. Z. Zmudzka, N. Kollias, and J. Z. Beer. In vivo measurement of skin erythema and pigmentation: new means of implementation of diffuse reflectance spectroscopy with a commercial instrument. *British Journal of Dermatology*, 159:683–690, 2008.

[58] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot. Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Trans. Geos. and Remote Sens.*, 47(8):2973–2987, 2009.

[59] J.C. Tilton. Image segmentation by region growing and spectral clustering with a natural convergence criterion. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 4, 1998.

[60] S. Valero, P. Salembier, and J. Chanussot. New hyperspectral data representation using binary partition tree. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 80–83, 2010.

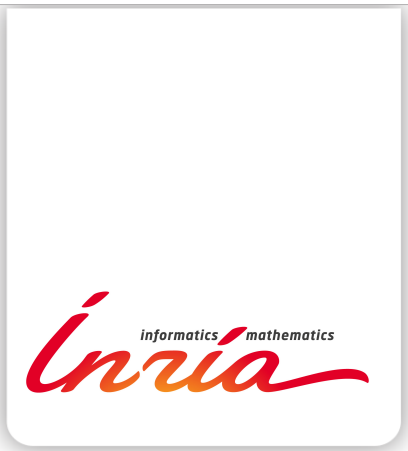
[61] S. Van der Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert. Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines. *Journal of Applied Remote Sensing*, 1, 2007.

[62] V. Vapnik. Statistical learning theory. *John Wiley and sons, inc.*, 1998.

[63] M. Winter. Fast autonomous spectral end-member determination in hyperspectral data. In *Conf. on Applied Geologic Remote Sensing*, volume 2, pages 337–344, 1999.

[64] Y. Yakimovsky. Boundary and object detection in real world images. *J. ACM*, 23:599–618, Oct. 1976.

[65] Z. Zhang and X Huang. Object-oriented subspace analysis for airbone hyperspectral remote sensing imagery. *Neurocomputing*, 73:927–936, 2010.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399